1 Necessary and sufficient conditions for exactness of SC-bound

In this chapter we study necessary and sufficient conditions for exactness of SC-bound. To formulate our main result we have to first introduce the concept of boolean cube.

Let's fix a set of classifiers A. Consider a classifier $a \in A$ with protective set X_a. Then the *boolean cube generated by* a is defined as a set of all classifiers whose error vectors differ from a only on objects from X_a:

$$B(a) = \Big\{a' \in \{0,1\}^L \colon n(a,x) = n(a',x)\ \forall x \notin X_a\Big\}.$$

Boolean cube $B(a)$ consists of $2^{|X_a|}$ classifiers identical to a in $\mathbb{X}\backslash X_a$ and having all feasible combinations of errors in X_a.

Theorem 1 *Let A be a set of classifiers and suppose that learning algorithm μ is a PessERM. Then SC-bound for A is exact whenever both of the following conditions are satisfied simultaneously:*

- *there is exactly one classifier $a \in A$ with $d(a) = 0$ (source), and exactly one classifier $b \in A$ with $u(b) = 0$ (sink);*
- *together with each $a \in A$ the set A contains boolean cube $B(a)$.*

It will be also proven that under some additional restrictions the statement of theorem 1 gives necessary and sufficient conditions for exactness of SC-bound.

We start the proof with the following simple lemma. Note that in this lemma and further in this chapter by X_a and X'_a we denote protective and prohibitive subsets for a, defined in the same way as in SC-bound:

$$\begin{aligned} X_a &= \big\{x_{ab} \in \mathbb{X} \bigm| a \prec b\big\} \ \text{ is the protective subset;} \\ X'_a &= \big\{x \in \mathbb{X} \bigm| \exists b \in A \colon b \leqslant a,\ I(b,x) < I(a,x)\big\} \ \text{ is the prohibitive subset.} \end{aligned}$$

Lemma 2 *Let A be a set of classifiers and suppose that learning algorithm μ is a PessERM. Then for any $\varepsilon \in (0,1)$ the SC-bound of A is exact iff for each $a \in A$ and for each $X \in [\mathbb{X}]^\ell$ at least one of the following two conditions is satisfied:*

1) $\big[\mu X = a\big] = \big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]$;
2) $\nu(a,\bar{X}) - \nu(a,X) < \varepsilon$.

Proof: Let's revisit the proof of SC-bound:

$$\begin{aligned} Q_\varepsilon &= \mathsf{P}\big[\mu X = a\big]\big[\nu(a,\bar{X}) - \nu(a,X) \geqslant \varepsilon\big] \leqslant \\ &\leqslant \mathsf{P}\big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]\big[\nu(a,\bar{X}) - \nu(a,X) \geqslant \varepsilon\big] = \tilde{Q}_\varepsilon. \end{aligned}$$

Then the difference $\tilde{Q}_\varepsilon - Q_\varepsilon$ can be written as a sum of non-negative summands:

$$\tilde{Q}_\varepsilon - Q_\varepsilon = \sum_{X \in [\mathbb{X}]^\ell} \Big(\big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big] - \big[\mu X = a\big]\Big) \cdot \big[\nu(a,\bar{X}) - \nu(a,X) \geqslant \varepsilon\big].$$

Therefore, SC-bound is exact iff each summand is zero, which is equivalent to the statement of the lemma. ∎

From now we will focus on the first condition of lemma 2. If first condition is satisfied for all $a \in A$ then SC-bound is exact. The main advantage of this condition is that it only relies on the structure of the set A, but doesn't depend on ε.

Lemma 3 *Let A be a set of classifiers and suppose that learning algorithm μ is a PessERM. Consider arbitrary $a \in A$. Then equality $\big[\mu X = a\big] = \big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]$ holds for all $X \in [\mathbb{X}]^\ell$ whenever for each $b \in A$, $b \neq a$ at least one of the following two conditions is satisfied:*

$$\big(X_a \cap X'_b \neq \varnothing\big) \text{ or } \big(X'_a \cap X_b \neq \varnothing\big). \tag{1}$$

If, in addition, the length parameters of train and test subsamples satisfy $\ell \geq 2 \max\limits_{a \in A} u(a)$ and $k \geq 2 \max\limits_{a \in A} q(a)$, then (1) also gives necessary condition for $\big[\mu X = a\big] = \big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]$.

Proof: Let $X \in [\mathbb{X}]^\ell$ be an arbitrary partitioning of $\mathbb{X}$ into train and test subsamples X and $\bar{X}$. Let $a = \mu X$. To prove the lemma we have to consider arbitrary $b \in A$, $b \neq a$, and prove the following:

$$\big[X_b \subset X\big]\big[X'_b \subset \bar{X}\big] = 0.$$

From $\mu X = a$ we conclude that $X_a \subset X$ and $X'_a \subset \bar{X}$. From the conditions of the lemma either $X_a \cap X'_b \neq \varnothing$ or $X'_a \cap X_b \neq \varnothing$. If the former is true than there is an object from X'_b belonging to X_a, and hence belonging to X. Therefore $[X'_b \subset \bar{X}] = 0$. If the later is true than there is an object from X_b belonging to $\bar{X}$, therefore $[X_b \subset \bar{X}] = 0$. In both cases $\big[X_b \subset X\big]\big[X'_b \subset \bar{X}\big] = 0$, which proves the first part of the lemma.

To prove the second part of the lemma let's assume the opposite and consider a pair of different a, b from A with $X_a \cap X'_b = \varnothing$ and $X'_a \cap X_b = \varnothing$. Then $(X_a \cup X_b) \cap (X'_a \cup X'_b) = \varnothing$, and from the conditions of lemma $|X_a \cup X_b| \leq \ell, |X'_a \cup X'_b| \leq k$. Then there is a partition of $\mathbb{X}$ into train set X and test set $\bar{X}$ with $X_a \cup X_b \subset X, X'_a \cup X'_b \subset \bar{X}$. But this implies $\big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big] = \big[X_b \subset X\big]\big[X'_b \subset \bar{X}\big] = 1$. This contradiction proves the second part of the lemma. ■

The next lemma studies properties of a set A when it is closed with respect to completion by boolean cubes.

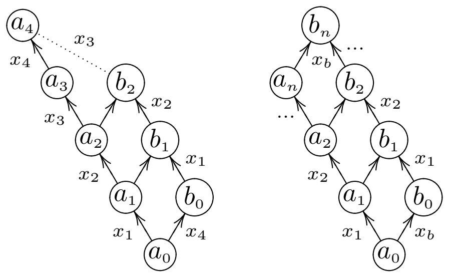
A *chain* $a_0 \to a_n$ from classifier a_0 to a_n is defined as a sequences $a_0 \prec a_1 \prec \cdots \prec a_n$. We say that n is the *length* of this chain.

Lemma 4 *Suppose that set A together with each classifier $a \in A$ contains boolean cube $B(a)$. Consider two classifiers a_0 and a_n such that there is a chain $a_0 \prec a_1 \prec \cdots \prec a_n$ belonging to A, and some classifier b_0 such that $a_0 \prec b_0$. Let $X'_i \equiv X'_{a_i}$ be a prohibitive set for a_i, and, for $i > 0$, let $x_i = X'_i \backslash X'_{i-1}$ be an object that differentiate a_i from a_{i-1}, $x_b = X'_{b_0} \backslash X'_0$ be an object that differentiate b_0 from a_0. Then*

1. *if $x_b \in \{x_1, \ldots, x_n\}$, then there exist a chain from b_0 to a_n;*
2. *if $x_b \notin \{x_1, \ldots, x_n\}$, then x_b must be in protective set for a_n.*

Proof: Consider first case when $x_b = x_p$ for some $p \in 1, \ldots, n$ (this is illustrated on the figure below, for $p = 4$). Note that both x_1 and x_b belong to protective set of a_0, therefore set A contains certain classifier b_1 such that $b_0 \prec b_1$ and $a_1 \prec b_1$. This argument can be iterated to deduce classifiers $b_2, \ldots, b_{p-2}$. Then one can easily see that $b_{p-2} \prec a_p$, because they differ by object x_{p-1}. Hence we have built the chain $b_0 \prec b_1 \prec \cdots \prec b_{p-2} \prec a_p \cdots \prec a_n$.

The case when $x_b \notin \{x_1, \ldots, x_n\}$ is also illustrated below. Then one can iterate the same arguments as before to build classifier b_n, such that $a_n \prec b_n$. This proves that x_b is in protective set for a_n.



■

We are now finally ready to prove our two main theorems.

Theorem 5 *Let A be a set of classifiers and suppose that learning algorithm μ is a PessERM. Then equality $\big[\mu X = a\big] = \big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]$ holds for all $X \in [\mathbb{X}]^\ell$ whenever both of the following conditions are satisfied simultaneously:*

1. *there is exactly one classifier $a \in A$ with $d(a) = 0$ (source), and exactly one classifier $b \in A$ with $u(b) = 0$ (sink);*
2. *together with each $a \in A$ the set A contains boolean cube $B(a)$.*

Proof: To conduct the proof we assume that set A has exactly one source and sink, and together with each $a \in A$ it contains boolean cube $B(a)$. We will now utilize lemma 3, and show that for every $a \in A$ and $b \in A, b \neq a$ either $X_a \cap X'_b \neq \varnothing$, or $X'_a \cap X_b \neq \varnothing$. Given that last two conditions are symmetric with respect to swapping $a \leftrightarrow b$, we will assume that $n(a, \mathbb{X}) \leqslant n(b, \mathbb{X})$.

Let's build some chains $s \prec a_1 \prec \cdots \prec a_n \equiv a$ and $s \prec b_1 \prec \cdots \prec b_p \equiv b$ from global source $s \in A$ to classifiers a and b. Here n is the length of chain $s \to a$, and p is the length of the chain $s \to b$. These chains exist due to uniqueness of source s (indeed, one may start from a or b and move downwards in SC-graph until he reaches source s). Denote objects, corresponding to these chains, by $X'_a = \{x^a_1, \dots, x^a_n\}$ and $X'_b = \{x^b_1, \dots, x^b_p\}$ (in the same order as classifiers from the chains make errors on these objects).

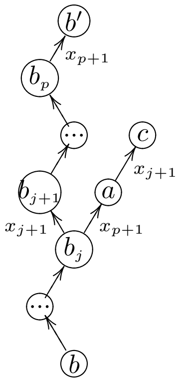
Let's select the smallest $t \in 1, \dots, p$ such that $x^b_t \notin X'_a$, and prove that object x^b_t belongs to X_a. This object exists either because $p > n$ and then the set X'_b is simply larger than X'_a, or, if $n = p$, because non-existance of such object would imply $X'_b = X'_a$, and therefore $a = b$ (which violates our assumptions).

Consider objects $x^b_1, \dots, x^b_{t-1}$. By definition of t they all belong to X'_a, therefore applying lemma 4 $t-1$ times we conclude that there exist a chain from classifier b_{t-1} to a. Now we again apply lemma 4 to chain $b_{q-1} \to a$ and classifier b_t. Since x^b_t doesn't occur in chain $b_{t-1} \to a$ we conclude that $x^b_t \in X_a$. As the result $X_a \cap X'_b \neq \varnothing$, and then by lemma 3 the theorem is proved. ■

Theorem 6 *Theorem 5 also gives necessary conditions for* $\big[\mu X = a\big] = \big[X_a \subset X\big]\big[X'_a \subset \bar{X}\big]$ *whenever length parameters* ℓ *and* k *of training and testing sample satisfy* $\ell \geq 2 \max\limits_{a \in A} u(a)$ *and* $k \geq 2 \max\limits_{a \in A} q(a)$.

Proof: We only study the case $|A| > 1$ (if $|A| = 1$ then the theorem is trivial). Note that there exist at least one source and one sink because any classifier from the lowest / the highest layer of SC-graph is a source / a sink. Therefore it is sufficient to prove that under conditions of this theorem they are unique. Consider the opposite, and let s_1 and s_2 be two sinks. Then $X_{s_1} = X_{s_1} = \varnothing$, and this violates conditions (1) of lemma 3. Note that the same argument fails to work for sources, because $X'(a) = \varnothing$ doesn't automatically follow from $d(a) = 0$. To prove that source is unique we take any classifier a from the lowest layer of A, consider any other $b \in A$ such that $d(b) = 0$ and $b \neq a$, and show contradiction. Indeed, if $X'(b) = \varnothing$ then we've again got a contradiction with (1) of lemma 3. Otherwise, if $X'(b) \neq \varnothing$, there exist some $c \in A$, $c < b$. Due to lemma 3 we have $X_c \cap X'_b \neq \varnothing$, hence there exist certain classifier c' such that $c \prec c'$ and still $c' < b$. By applying last arguments several times we build a chain $c \prec c' \prec \cdots \prec b$, causing b to have an incoming edge in SC-graph, and therefore contradicting $d(b) = 0$.

Now we turn to the second part and prove that together with each a the set A contains boolean cube $B(a)$. Suppose the opposite, then there exists a classifier $b \in A$ such that $B(b) \nsubseteq A$, i.e. A doesn't fully contain boolean cube generated by object b. Let b' be the classifier from $B(b) \setminus A$ with minimal number of errors. Since all classifiers from $B(b)$ whose number of errors smaller than $m(b')$ are contained in A, we can construct a chain $b \equiv b_0 \prec b_1 \prec \ldots \prec b_p \prec b_{p+1} \equiv b'$ such that all classifiers except the last lie in A. For every $i = 1, \dots, (p+1)$ denote by x_i an object that corresponds to the edge between b_{i-1} and b_i. We also denote protective and prohibitive sets of $b_0, b_1, \dots, b_{p+1}$ by $X_0, X_1, \dots, X_{p+1}$ and $X'_0, X'_1, \dots, X'_{p+1}$ respectively.

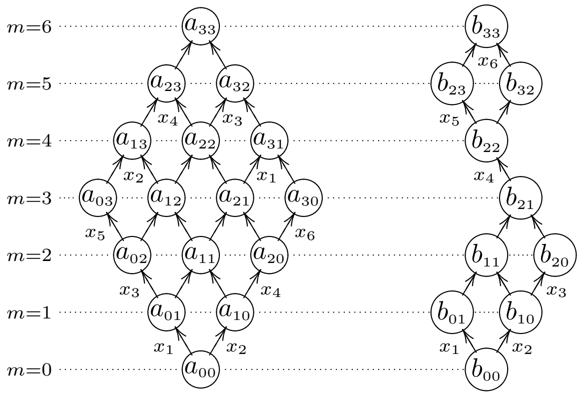


Note that $x_{p+1} \notin X_p$ because $b_{p+1} \notin A$. Now consider all classifiers b_i from the constructed chain such that $x_{p+1} \in X_i$. Note that there is at least one such classifier because $x_{p+1} \in X_0$ by definition of boolean cube generated by b_0. Among those classifiers we choose the one with the largest number of errors, and denote it by b_j.

Since x_{p+1} is contained in X_j then there is such $a \in A$ that $b_j \prec a$ (see a figure above). Note that sets X'_{j+1} and X'_a differ only in two objects: X'_{j+1} contains x_{j+1}, and X'_a contains x_{p+1} instead (it follows from the fact that $|X'_a| = |X'_{j+1}| = |X'_j| + 1$). Now, applying lemma 3 to classifiers b_{j+1} and a, either $X_{j+1} \cap X'_a \neq \varnothing$ or $X'_{j+1} \cap X_a \neq \varnothing$ must be true. These statements are equivalent to $x_{p+1} \in X_{j+1}$ and $x_{j+1} \in X_a$ accordingly. First suppose that $x_{p+1} \in X_{j+1}$; this is impossible because b_j by our choice has the largest number of errors among all classifiers whose protective sets contain x_{p+1}. Now suppose that $x_{j+1} \in X_a$; then there exists a classifier $c \in A$ with the same error vector as b_j except that $I(c, x_{j+1}) = 1$ and $I(c, x_{p+1}) = 1$ (again, see the figure above). But b_{j+1} also has the same error vector as b_j except that $I(b_{j+1}, x_{j+1}) = 1$. This means that b_{j+1} and c are connected by an edge in SC-graph that corresponds to x_{p+1}. Then $x_{p+1} \in X_{j+1}$, which again contradicts to our choice of b_j.

These two contradictions conclude the proof of the theorem. ■

Finally we present two examples of SC-graphs that satisfy conditions of theorem 1 and therefore have an exact SC-bound.



References